



携程技术中心



IT大咖说  
知识分享平台

# 携程技术沙龙

## 高吞吐消息网关的探索与思考

分享人：刘惊惊



# 刘惊惊

- 唯品会业务架构部高级架构师
- 负责用户，营销，库存系统的架构设计

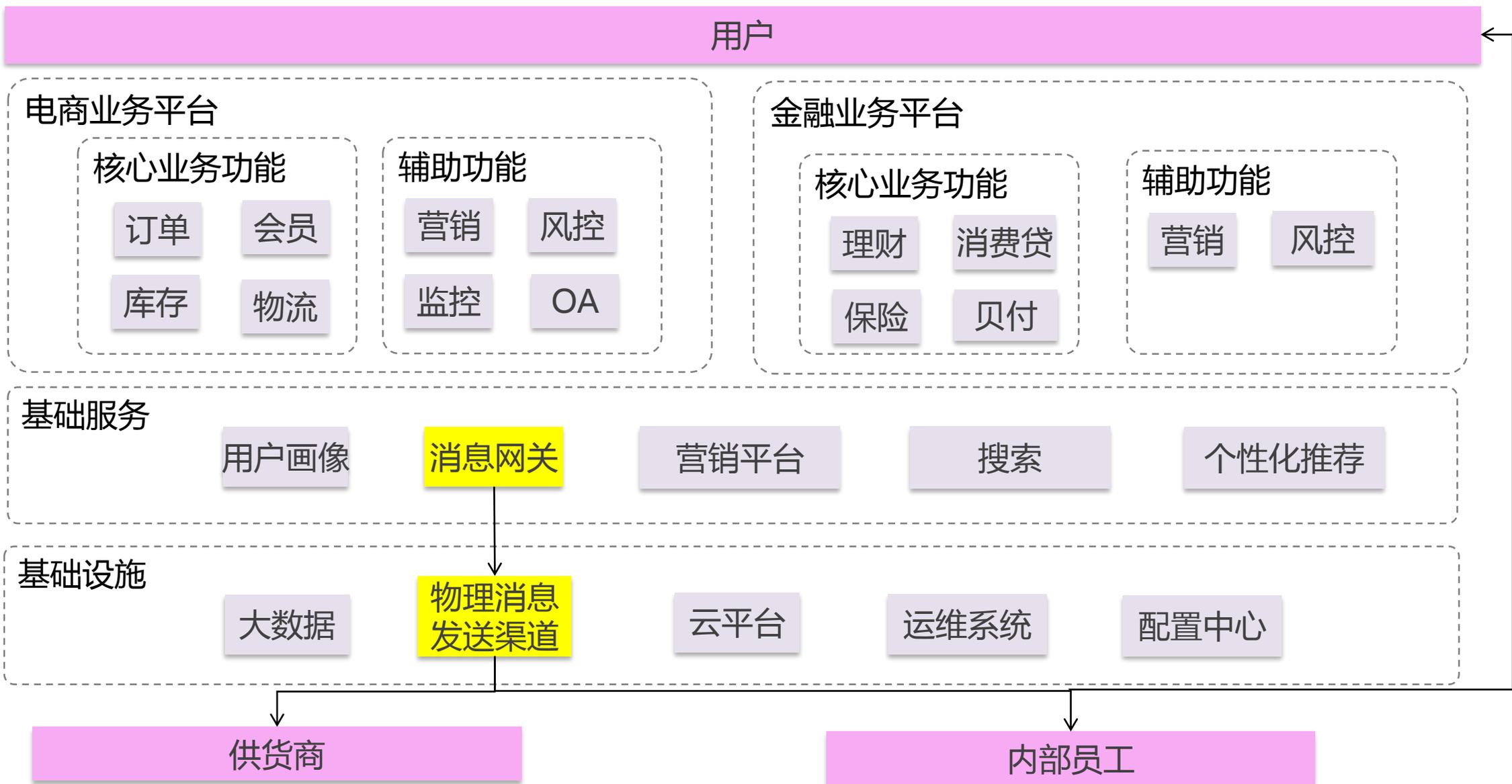
# 目录

CONTENTS

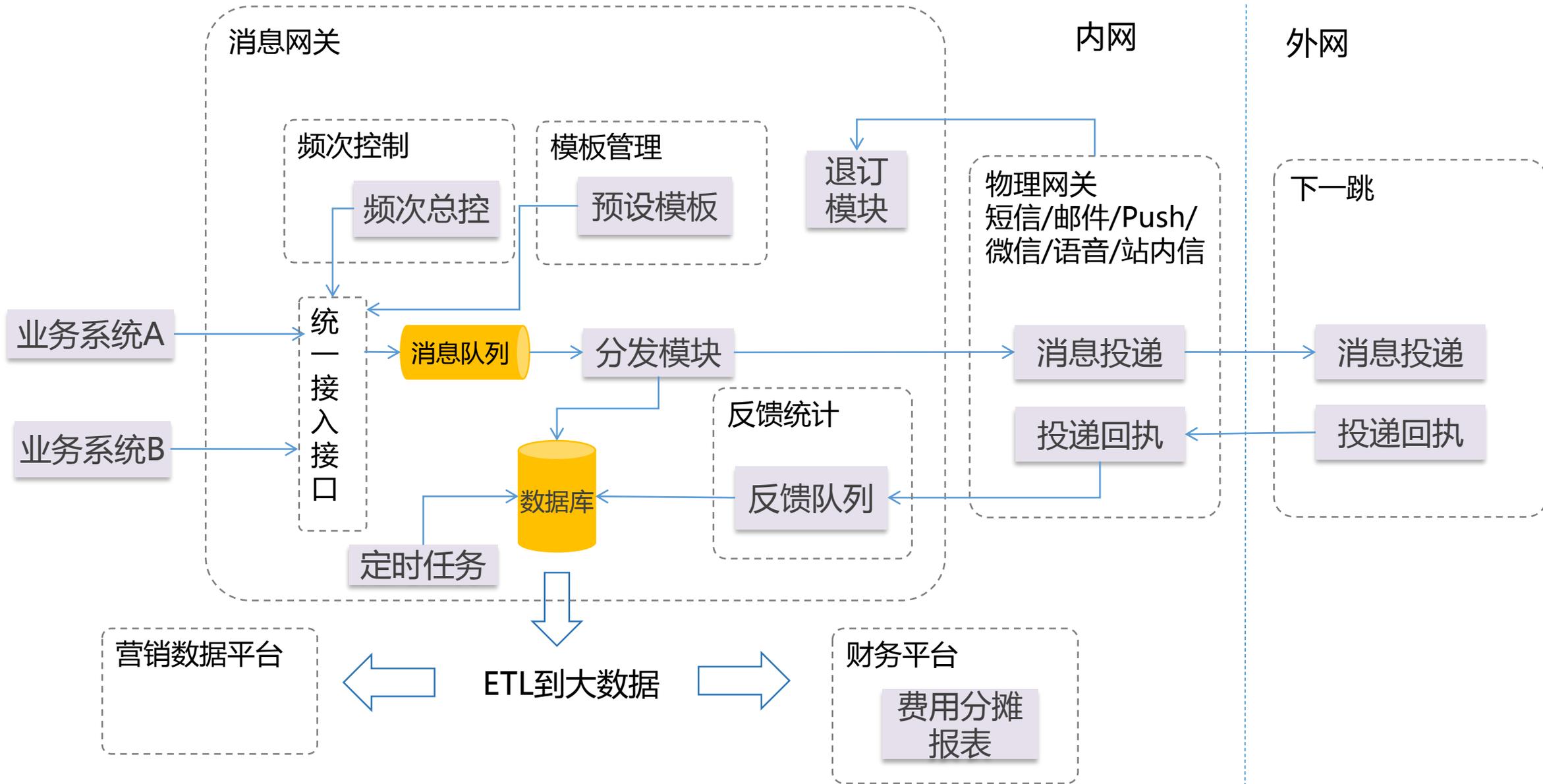
- 1 消息网关是什么
- 2 如何设计消息网关
- 3 消息网关的技术选型
- 4 监控与降级
- 5 关于消息网关的思考

# 消息网关是什么？

- 消息网关是唯品会触达用户的逻辑消息发送层。
- 消息网关的上游是各个业务平台，他们对消息网关的统一发送接口推送触达消息。
- 消息网关的下游是物理投递渠道，他们负责对接信道提供方（电信运营商，微信等）和具体发送。
- 消息网关的职责是受理投递、频次控制、尽力发送、反馈统计。
- 消息网关属于基础业务，不应该简单的划归到营销业务范畴，应该独立出来作为基础服务。
- 消息网关作为缓冲层，起到削峰作用，匹配物理消息网关的慢速发送。



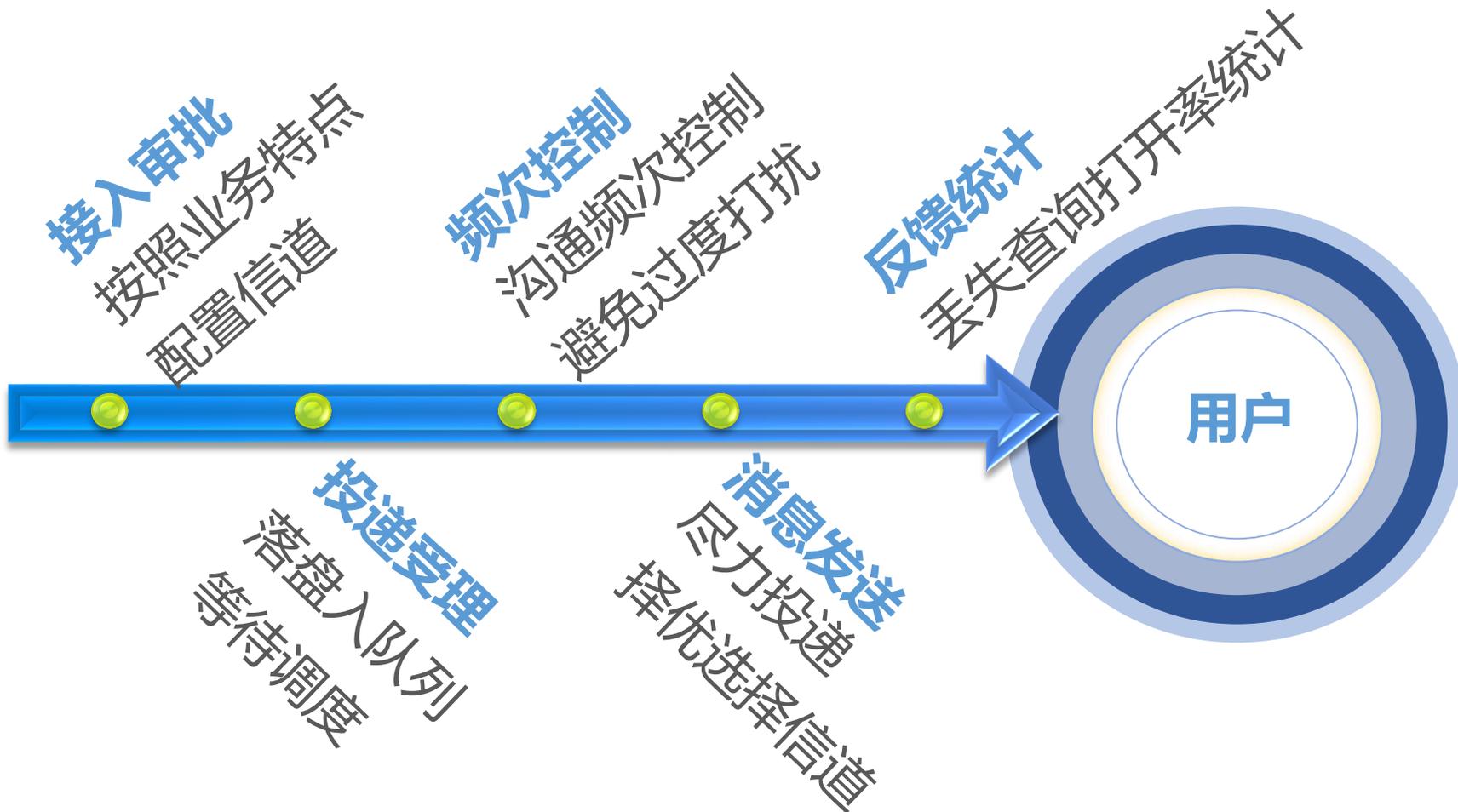
# 消息网关整体架构

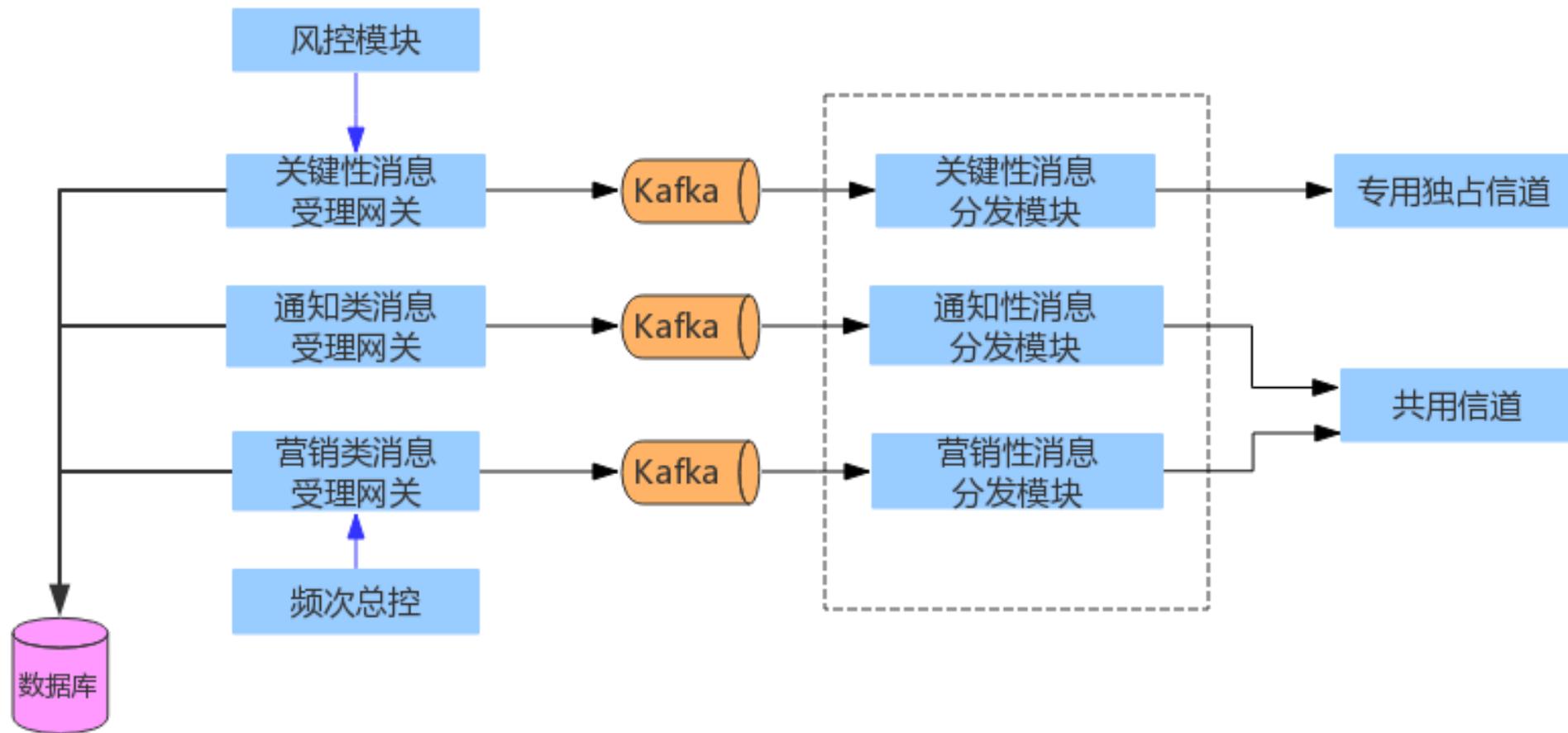


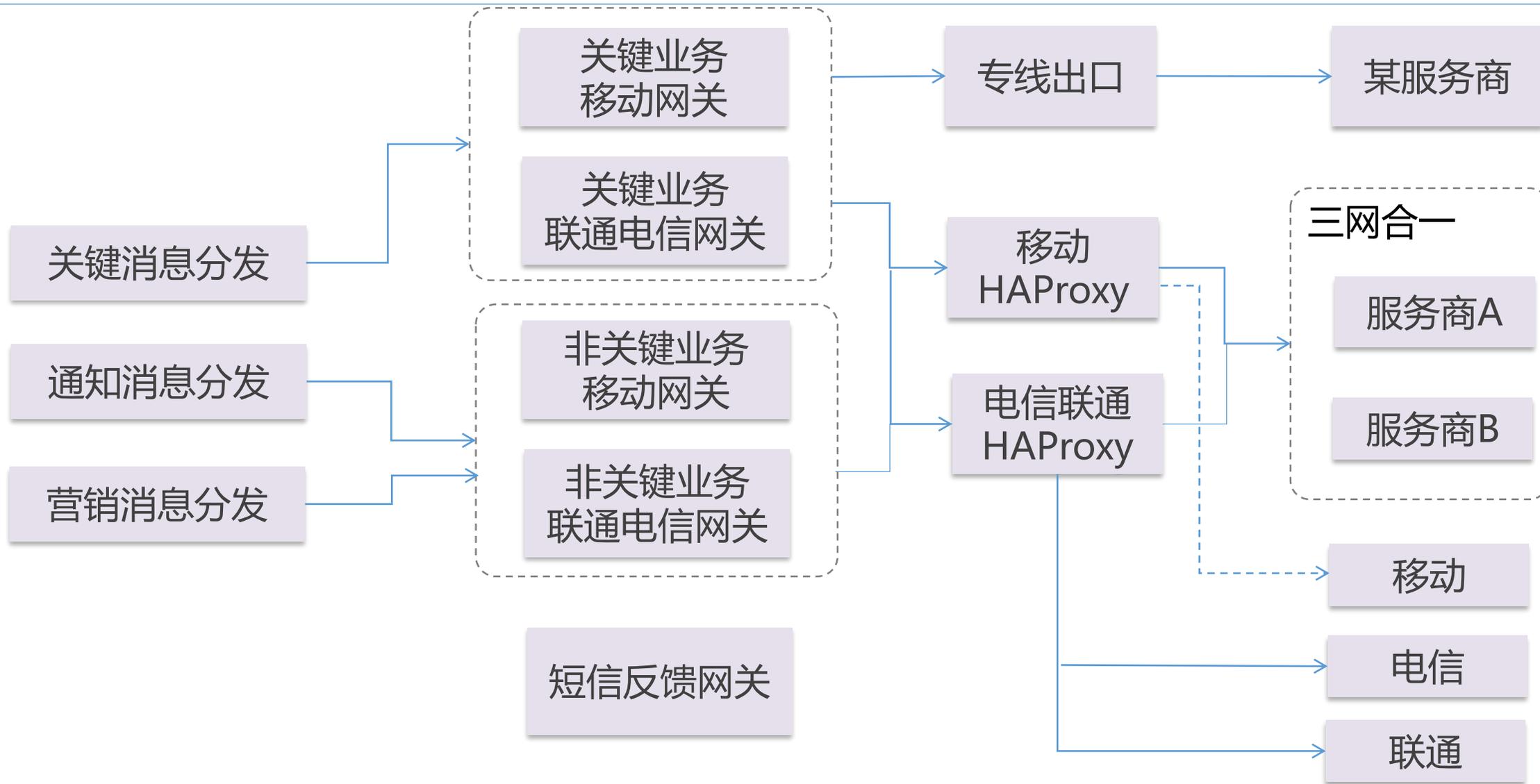
# 如何设计消息网关

- 时效性和吞吐量
- 提供业务方友好的接口
- 模板管理
- 频次控制
- 用户订阅
- 时效期内的失败重试
- 反馈统计
- 系统监控和降级
- 渠道动态规划

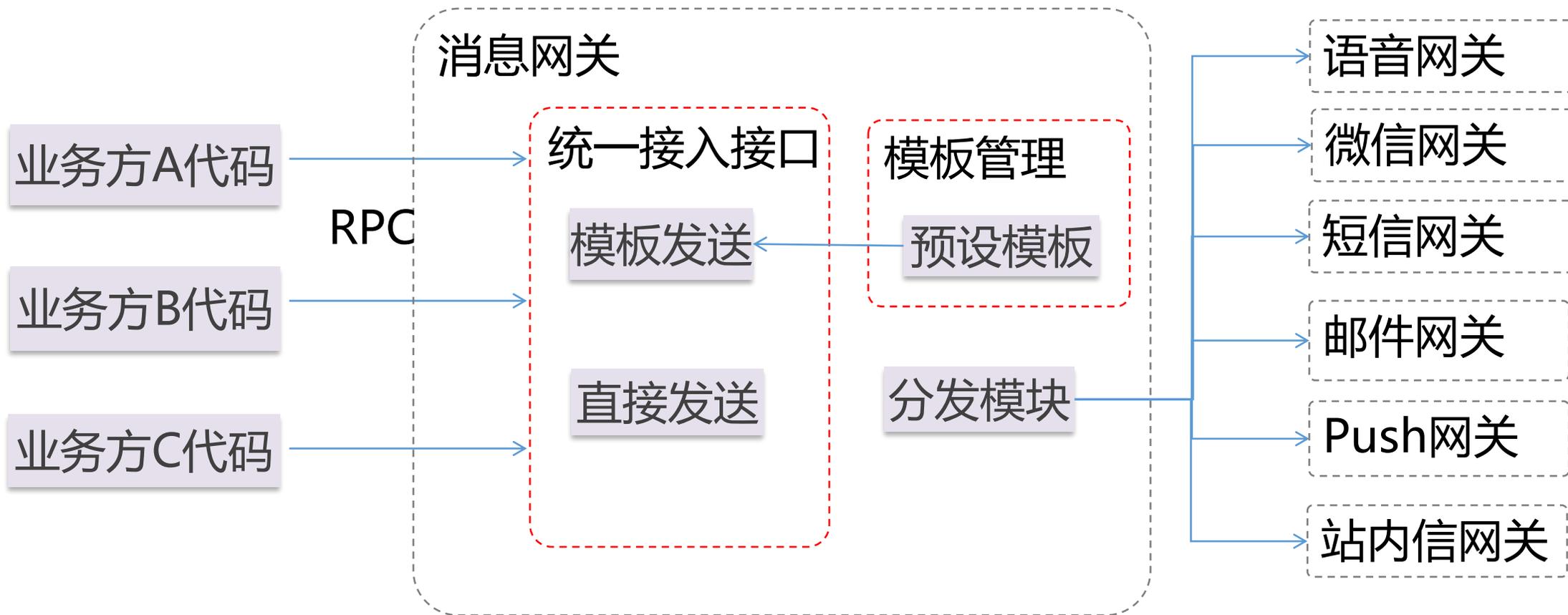
业务方接入

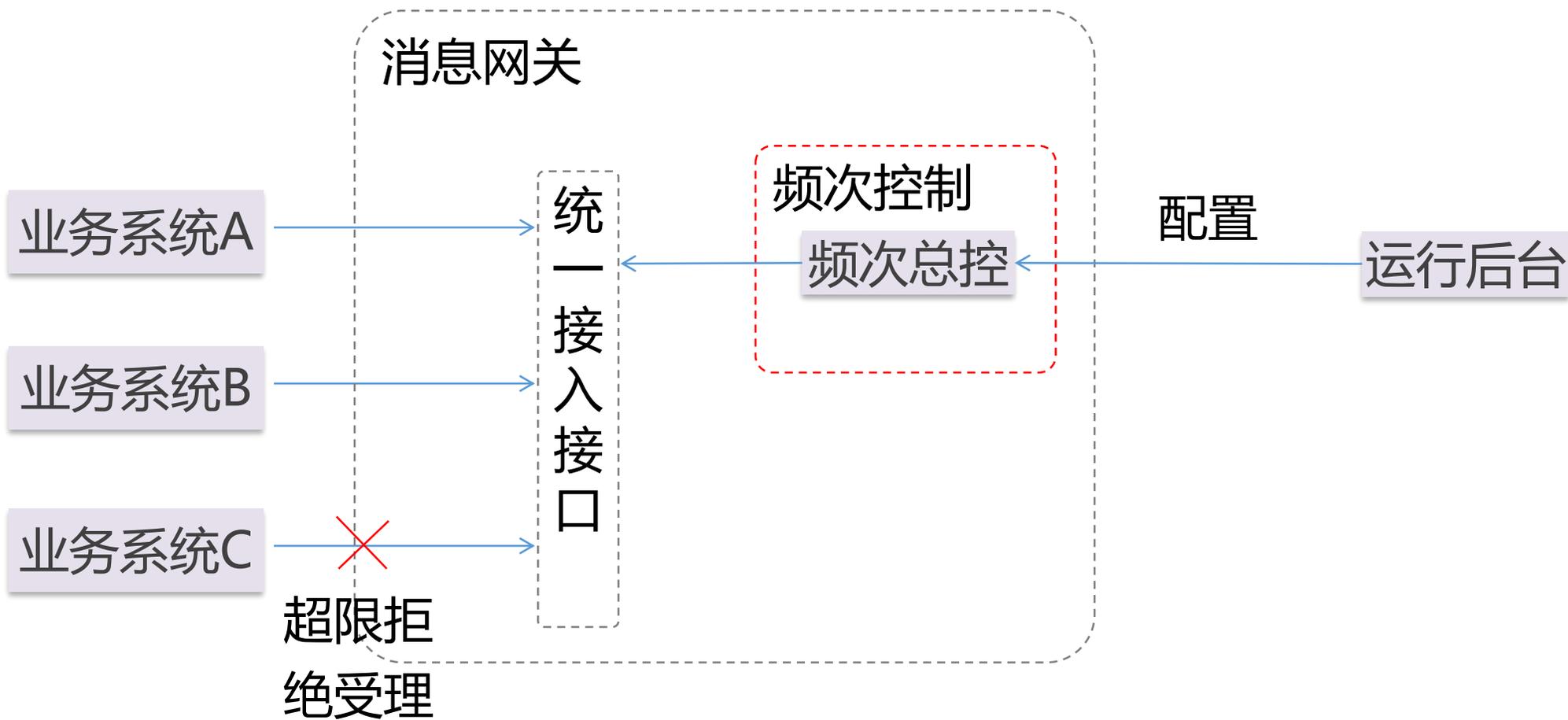




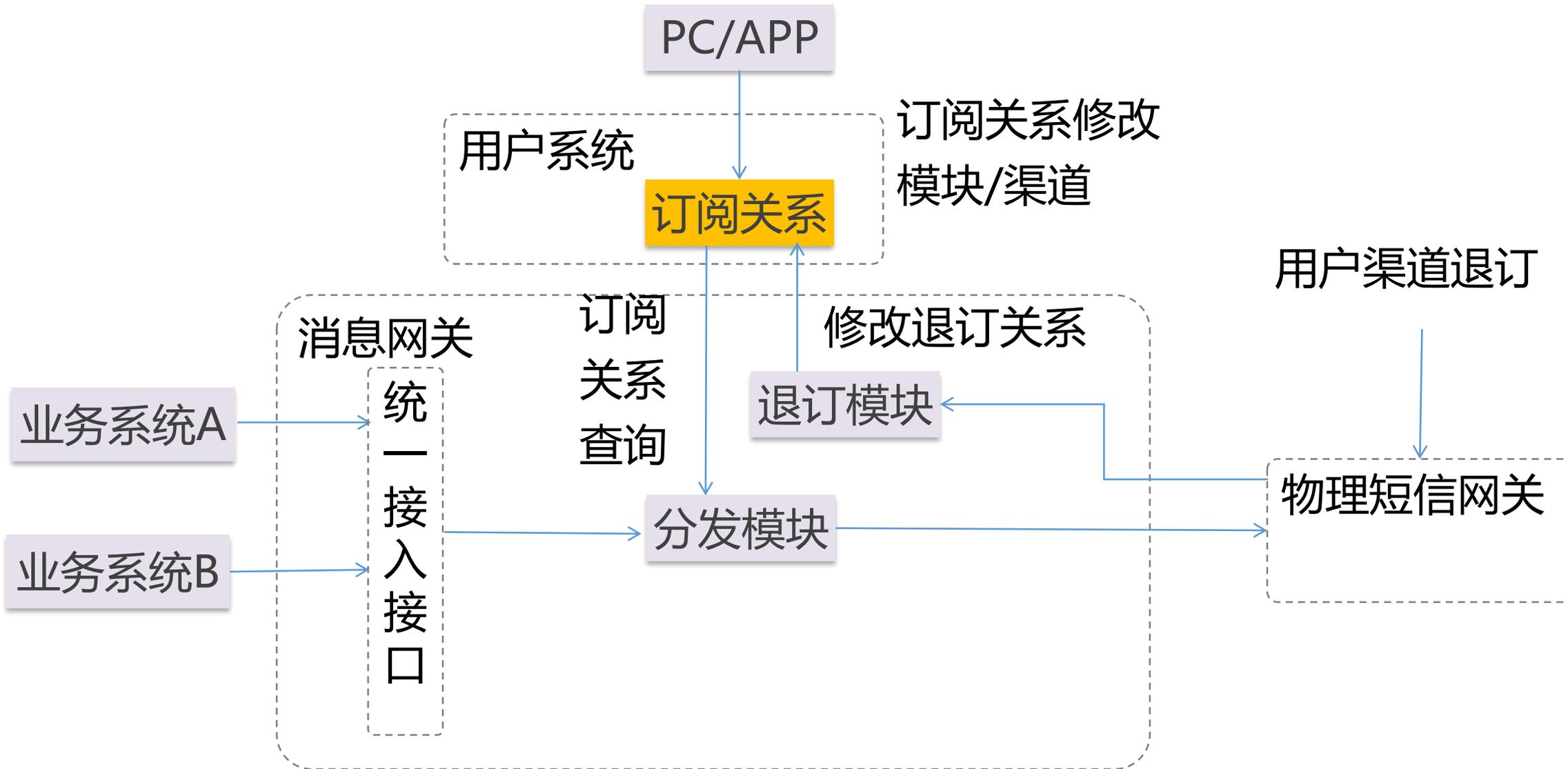


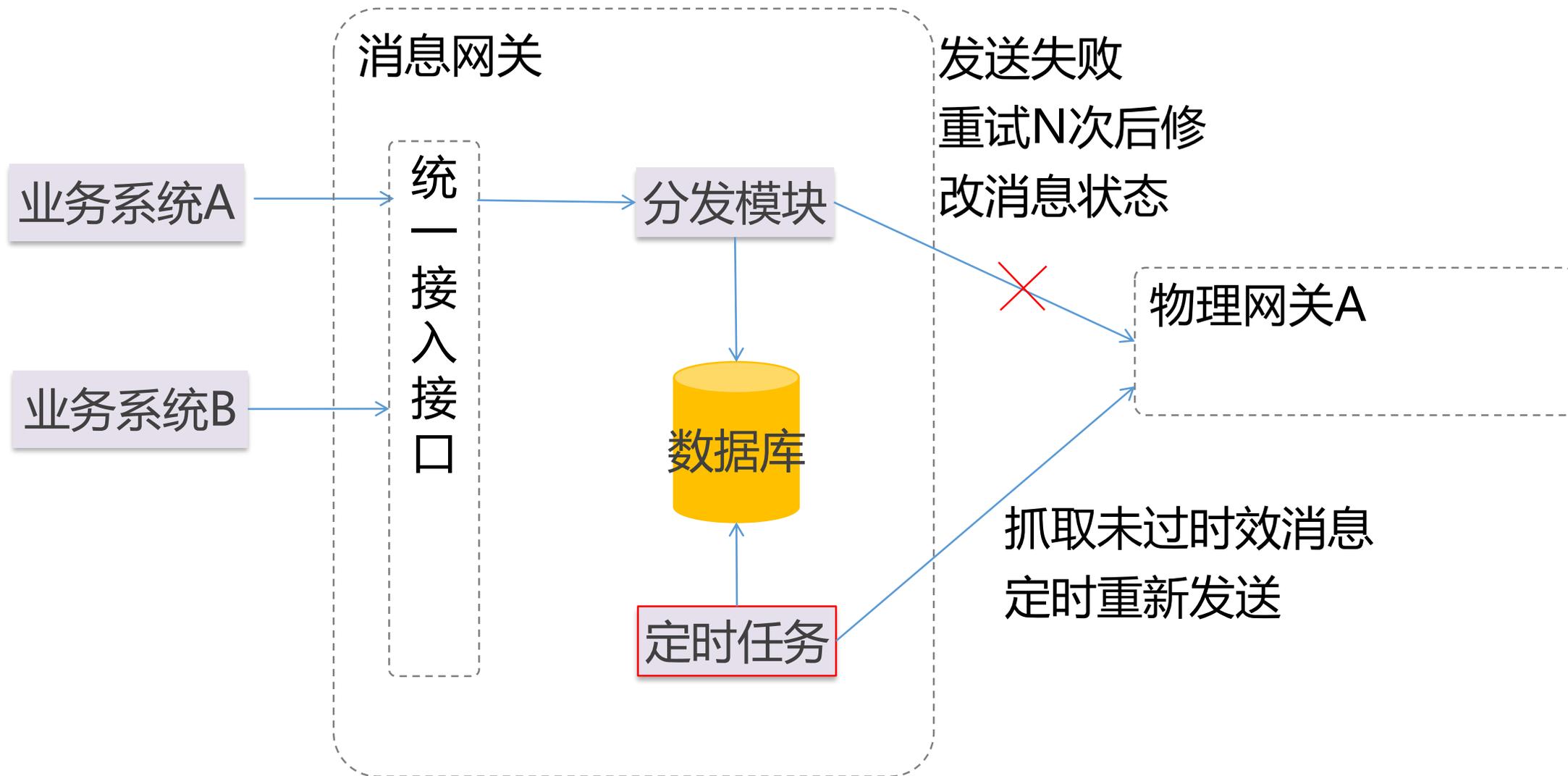
# 提供面对业务方友好的接口

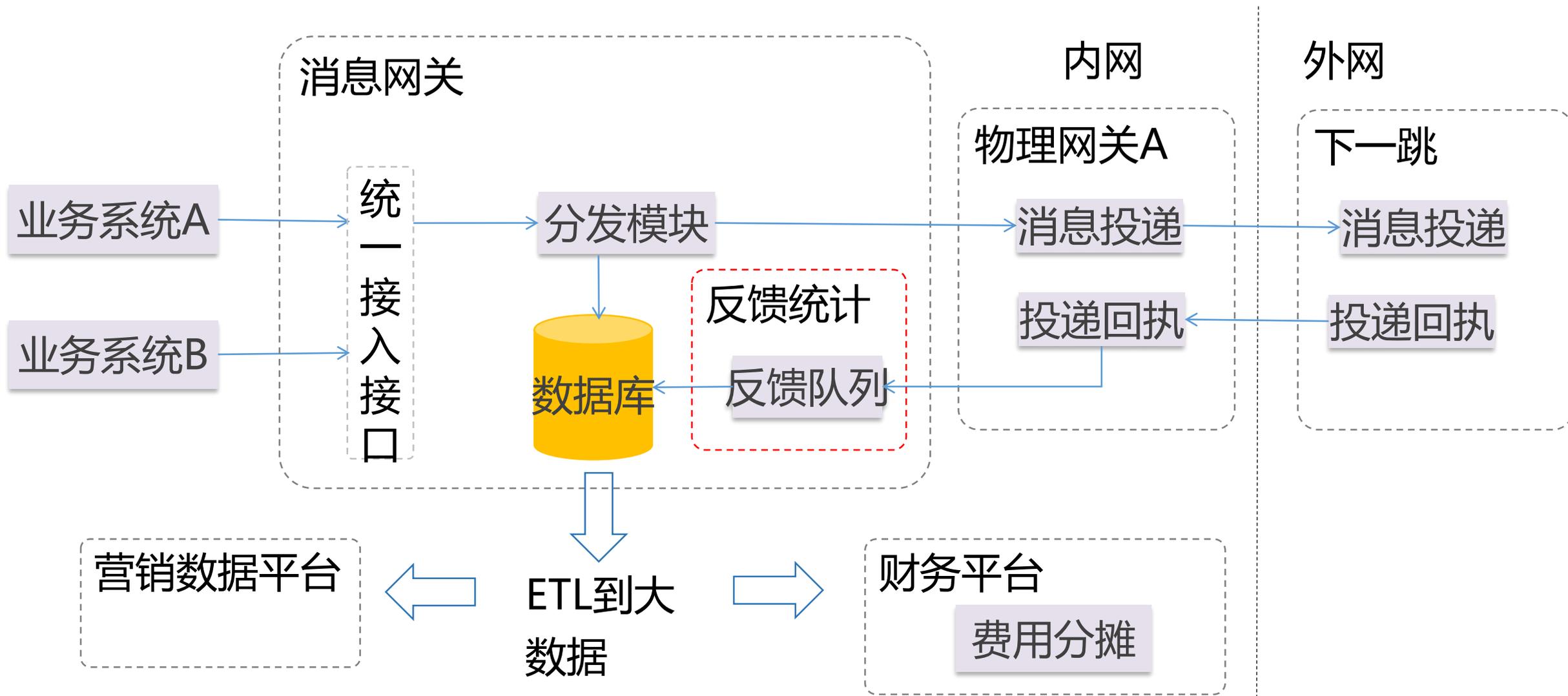




# 用户订阅

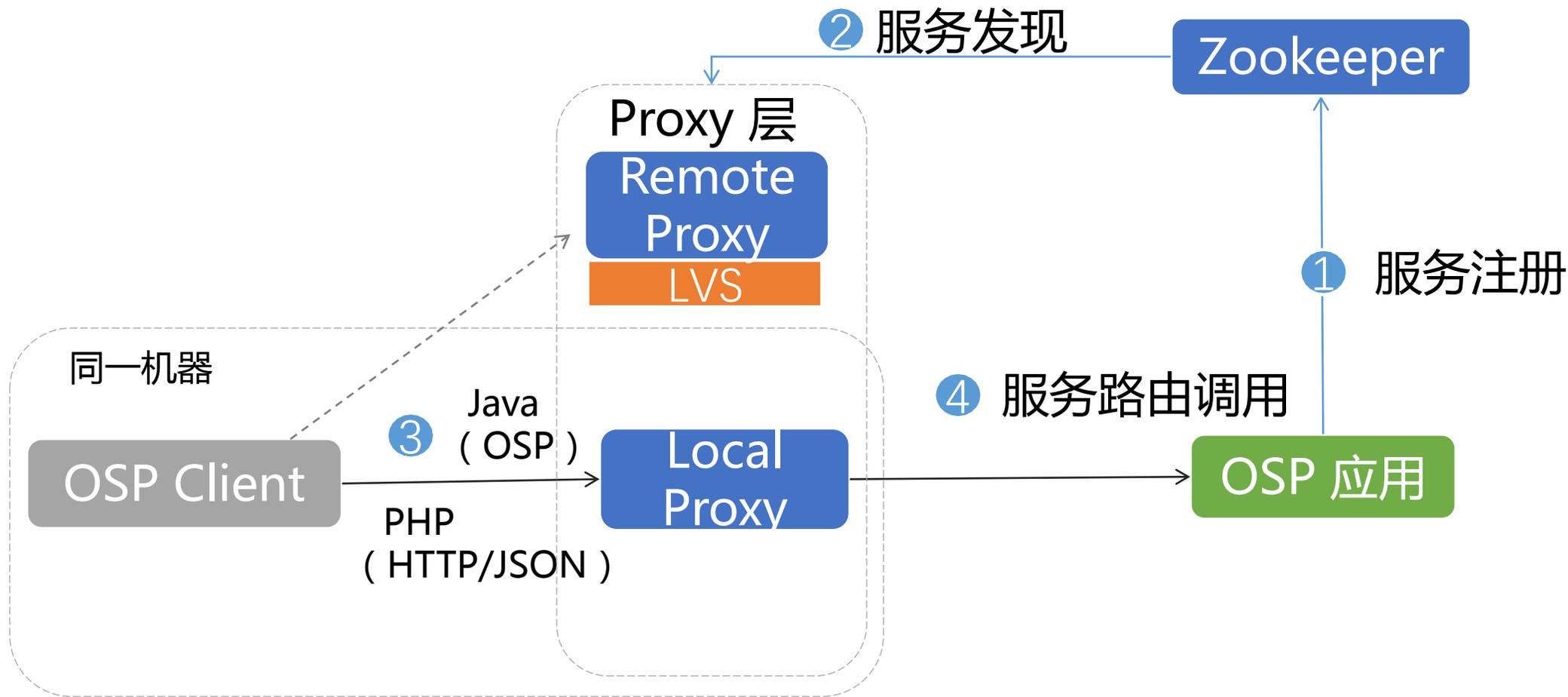






# 消息网关的技术选型

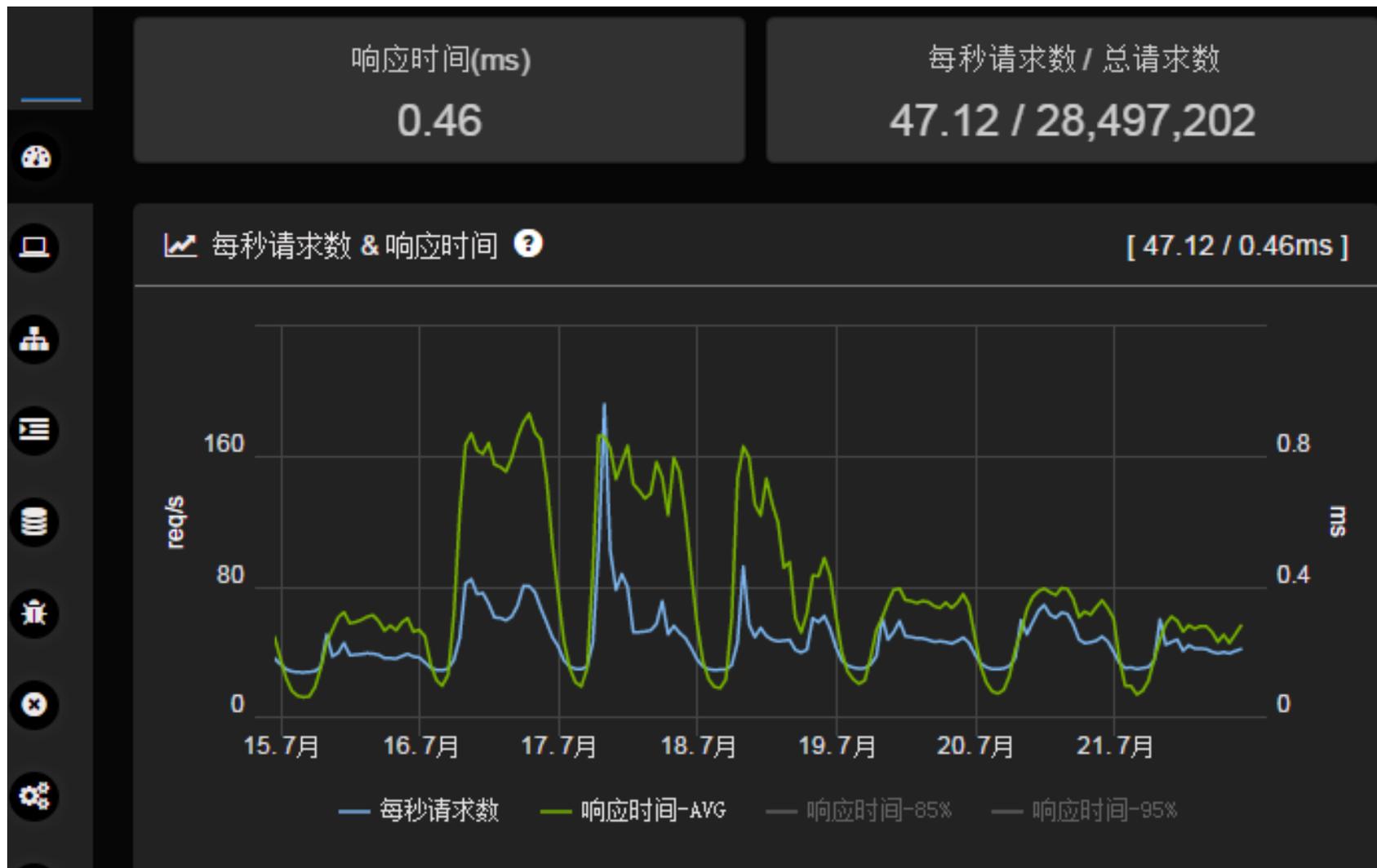
- 业务模块开发采用唯品会自研的Venus框架，采用OSP协议来封装RPC调用，底层通讯采用Netty，传输协议Thrift。
- 数据库技术选型MySQL。使用自研VDP(类似阿里Canal)，解析binlog，同步数据到HDFS。
- 消息队列用Kafka，应对高吞吐量消息场景。



# 监控与降级

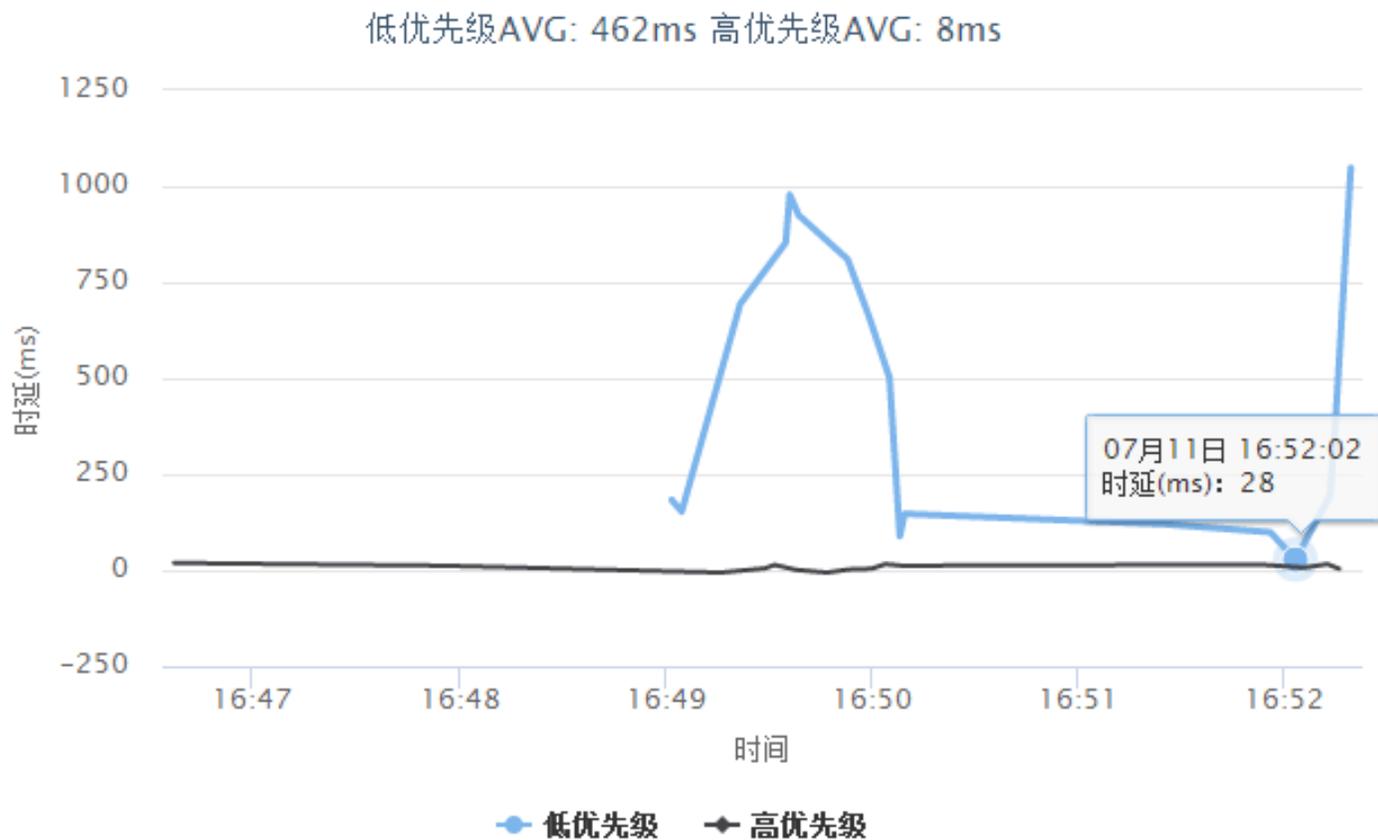
- 受理域，分发域各接口QPS监控（RPC框架内置）
- Kafka队列的堆积消息监控。
- Zabbix监控数据库CPU负载，内存使用量，磁盘IO
- 物理渠道投递时延监控。

# 受理域调用监控 - 重构后



## 时延监控

### 短信发送全链路时延监控



队列查询 > 队列详情

开始时间: 2017-06-15 19:58

结束时间: 2017-06-15 20:20

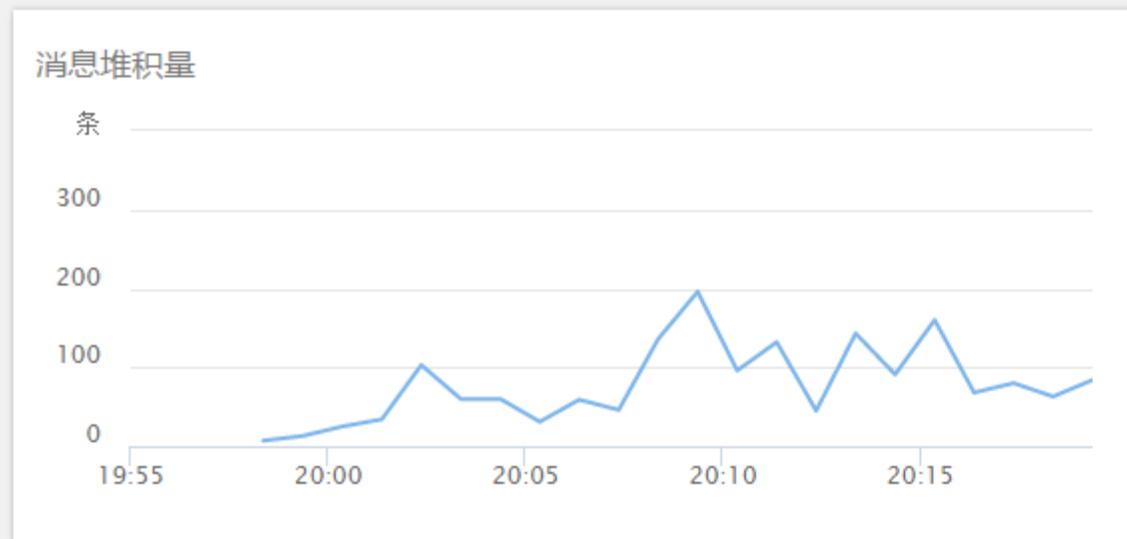
查询

当前队列: queue.umc.sms.app.highpriority

集群: vms.kafka.umc\_captcha

[查看绑定关系](#)

自动刷新: 关闭

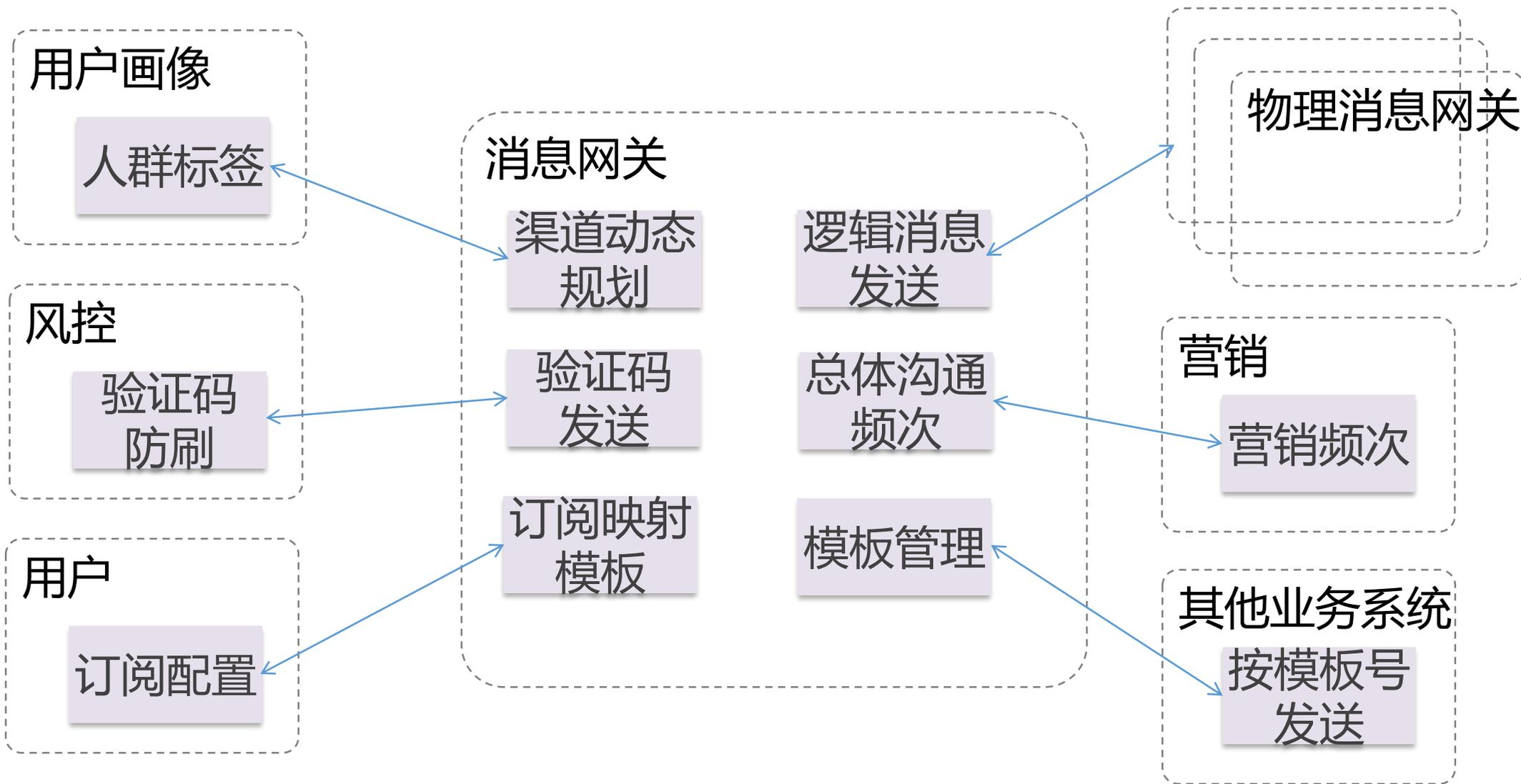


- 受理域机器宕机，由Local Proxy调度转发到正常工作的机器上。
- Kafka无法提供服务，切换到备用集群。已入队列未发送消息，从数据库中获取。
- 数据库主库宕机，影响消息状态变更和消息异步。切换到从库。不影响消息的正常发送，影响反馈和统计。
- 用户系统订阅关系无法返回结果，降级为不再查询，默认全部订阅。
- 物理消息网关宕机，投递失败，有重试机制保证。消息本身不丢失，可以在物理渠道恢复后，重发有效期内的消息。

2017.6.16大促支撑的消息峰值发送量。

|       | 通知/营销短信  | 支付类短信  | 邮件     | Push  | 站内信    |
|-------|----------|--------|--------|-------|--------|
| 峰值发送量 | 每分钟数十万量级 | 每分钟万量级 | 每天百万量级 | 每天亿级别 | 每天百万量级 |
|       |          |        |        |       |        |

# 关于消息网关的思考



- 营销消息的投递数据，是营销画像的重要组成部分
- 消息投递情况、打开率、花费影响到营销强度的计算
- 各个业务部门费用分摊的主要依据，设计上市公司财务审计
- 渠道服务提供商的选择和替换

- 短信网关的通讯协议CMPP/SMGP/SGIP，基于TCP协议上，需要自行处理断包粘包问题。配置到Mina框架NioSocketConnector的FilterChain里面，CmppDecoder（移动）/SmgpDecoder（电信）/SgipDecoder（联通）。用IoSession来完成收发操作。
- 异步引起的问题，消息数据异步落盘，可能落后于消息数据更新。引进延时队列，定时重试解决此问题。

- Redis存储Value的String不要超过2k字节（理论最大值512M），以不超过1个网络包(MTU 1500byte)为佳。Set和Sorted的elements不要超过5000个。

- 测试机和生产环境配置不同，压测导致NoHttpResponseException

/etc/sysctl.conf

net.ipv4.tcp\_syncookies = 1

net.ipv4.tcp\_tw\_reuse = 1

net.ipv4.tcp\_tw\_recycle = 1

net.ipv4.tcp\_fin\_timeout = 15



携程技术中心



IT大咖说  
知识分享平台

# THANK YOU!

---

## Q&A